DOCUMENT RESUME

ED 406 449 TM 026 431

AUTHOR	Taylor, Catherine S.
TITLE	The Influence of Task Directions on Student Performance for Open-Ended Mathematics Assessments.
PUB DATE	Apr 96
NOTE	29p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Dropouts; High Schools; *High School Students; *Mathematics Tests; *Performance Factors; Probability; Responses; *Scoring; *Test Items; Test Results
IDENTIFIERS	*Directions; Open Ended Questions

ABSTRACT

This study investigated the impact of task directions on the mathematical performance of high school students from six classes. Students analyzed data regarding school dropout by answering six short-answer questions and writing a letter discussing the trends and their predictions about school dropout. Tasks were scored using two methods: (1) trait scoring of students' responses for understanding statistical and probability concepts and for mathematical communication; and (2) item by item scoring rules. Trait scoring rules were applied by making holistic judgments about a student's collection of responses to a set of related items. For the item-by-item scoring procedure, scores for six short-answer items were summed for a statistics total score. The scores for students' letters were analyzed separately from the other six items. In one form, 140 students were told to add their own ideas to the letter about school drop-out (connections condition). There were 36 students in the no-connections condition. For trait scoring, there were no significant differences between groups for either the conceptual or communication traits. For item-by-item scoring, mean scores for the statistics total score were not different. However, letter scores for students in the "no connections" condition were significantly higher than letter scores for the students in the "connections" condition. Results are presented, and implications for test development and mathematics instruction are discussed. (Contains 4 figures, 9 tables, and 12 references.) (Author/SLD)

The influence of task directions on student performance for open-ended mathematics assessments

Catherine S. Taylor

University of Washington

Submitted: February 5, 1996

Educational Psychology

312 Miller Hall, Box 353600

University of Washington

Seattle, WA 98195-3600

Running Head: Performance on Mathematics Tasks

Paper Presented at the 1996 annual meeting of the National Council on Measurement in Education, New York.

Abstract

This study investigated the impact of task directions on students' mathematical performance. Students analyzed data regarding school dropout by answering six short-answer questions and writing a letter discussing the trends and their predictions about school dropout. Tasks were scored using two methods: a) trait scoring of students' response for understanding of statistical and probability concepts and for mathematical communication and b) item by item scoring rules. Trait scoring rules were applied by making holistic judgments about a student's collection of responses to a set of related items. For the item-by-item scoring procedure, scores for six short answer items were summed for a statistics total score. The scores for students' letters were analyzed separately from the other six items. In one form, students were also told to add their own ideas to the letter about school drop-out. For trait scoring, there were no significant differences between groups for either the conceptual or the communication traits. For item-by-item scoring, mean scores for the statistics total score were not different, however, letter scores for students in the "no connections" condition were significantly higher than letter scores for the students in the "connections" condition. Results are presented and implications for test development and mathematics instruction are discussed.

The influence of task directions on student performance for open-ended mathematics assessments

As more large-scale testing programs incorporate performance-based assessments, researchers are beginning to investigate issues relevant to these "alternative" assessment formats. Shavelson, and his colleagues (Shavelson, Baxter, & Pine, 1992; Shavelson, Baxter, 1992; Baxter, Shavelson, Herman, Brown, and Valdez, 1993) were among the first to move beyond rhetoric and consider factors that affect student score reliability for both mathematics and science assessment tasks. They found that examinee scores varied from task to task, indicating that performance-based assessments were highly instructionally sensitive and many such tasks would be needed to obtain reliable examinee scores. Goldberg and Kapinus (1993) closely examined responses to reading assessments and found that nuances in task directions could lead to differential interpretations. It appeared that examinees' construction of meaning included the meanings associated with the tasks themselves. The findings of these researchers should not be surprising, however. They confirm findings from writing research that peaked in the 1980s.

During the 1980's, when direct writing assessments were becoming standard practice in large-scale assessment programs, considerable research was conducted to examine issues that impacted the effectiveness of such programs. One area of research that received quite a bit of attention was the nature of writing prompts. Test developers and researchers were attempting to identify the characteristics of prompts that would optimize examinees' responses. For example, Cherry (1989) found that when examinees were unclear as to the rhetorical purpose of an essay, they had difficulty taking a position and defending it. Ruth and Murphy (1984) provided a model to explain possible sources of 'misfire' in writing assessments. In their model, they included examinee interpretations of writing topics, raters interpretations of writing topics, and raters interpretations of written essays. They noted that teachers could clarify directions in the classroom, however, in large scale assessments, "the individual problem of meaning [remains] unattended in the impersonal examination room." (p. 410)

In light of this problem of meaning construction, Hoetker, Brossell and Ash conducted a number of research studies to investigate variations in prompts (Brossell, 1983; Brossell & Ash, 1984; Hoetker, 1982; Hoetker & Brossell, 1986, 1989). Most of these studies showed that degree of specificity and rhetorical context had little effect on holistic essay scores.

To date, while no published studies have looked at variations in task directions for open-ended mathematics assessments, a few studies have focused on formats of assessment. For example, Shavelson, Baxter, and Pine (1992) compared hand-on tasks, computer simulation tasks, and paper-pencil tasks to see whether scores for tasks measuring the same science content were equivalent across methods. They found that scores were not consistent across task formats for the same science content. It is clear that similar studies are needed for large scale mathematics assessments. What is needed now are studies that investigate how variations in the assessment tasks given to students influence their performances.

The *Curriculum and Evaluation Standards for school Mathematics*, developed by the National Council for Teachers of Mathematics recommend that students be taught how to connect mathematical concepts and ideas across mathematical areas, between mathematics and other disciplines, and from mathematics to the real world. Although test developers are now working on developing tasks that can assess students' connections, little is really known about how to <u>assess</u> whether students can make these connections. In pilot studies conducted in the state of Washington, it was found that, although teachers could create contexts that modeled valid connections during the assessment process, it was difficult to identify ways to *ask students* about connections and actually elicit targeted learning.

The study presented here was designed to look at what happens when students are asked to make connections between their own ideas and the mathematical information presented in an assessment task. The goal was to determine whether directions to use prior knowledge, ideas, and opinions affected students' scores on mathematics tasks.

Methodology

The study was part of a pilot testing program for prototypic alternative assessments in the state of Washington. The purpose of the program was to develop models for assessments to help guide the development of the upcoming state assessment system. For the prototype pilot program as a whole, eight different task models were developed and piloted at each grade level in mathematics.

In order to locate pilot sites, administrators from throughout the state of Washington were invited to participate in a pilot testing program in the spring of 1995. The pilot testing program was described and contact persons were asked whether individuals in their district or school would be interested in participating in these pilots. If so, they were asked to volunteer 1 or 2 heterogeneously grouped classrooms at grades 4, 5, 7, 8, 10 and 11 for the pilots. Individuals from 33 of the state's 296 school districts volunteered. In a follow-up letter, volunteers were asked to identify teacher names and their content areas, the number of students in a class, and shipping addresses for materials. Once all pilot sites were identified, test forms were randomly assigned to classrooms.

This study focused only on two mathematics assessment forms at the high school level. A total of six classrooms were assigned each of the two mathematics forms. Materials were packaged for individual teachers and sent either to the district coordinator or the individual teachers, depending on the shipping address provided by the district coordinator. Materials included: a) a general overview of the pilot testing program with a description of the assessment types being piloted, b) oral directions along with specific directions for administering the writing tasks, c) sufficient student response books for one class of students, d) parent permission forms, e) student survey forms, and f) postage paid return envelopes.

Subjects

Table 1 describes the sample for this study. Materials were returned from three teachers for the "no connections" condition and five teachers for the "connections" condition. There

were 36 students in the "no connections" condition (all tenth graders) and 140 students in the "connections" condition (11 percent tenth graders and 89 percent eleventh graders).

---

insert Table 1 about here

Instrument

Assessment Tasks. The tasks were drafted by a pairs of teachers working on the mathematics assessment prototypes. Tasks were edited by a professional test development staff and reviewed by all item/task writers for mathematics. Tasks were then pre-piloted with 8 to 10 students to determine whether directions made sense. They were then revised based on the results of the pre-pilots and input from the reviewers. Finally, tasks were prepared for printing and distribution.

For the mathematics tasks, the teachers endeavored to create tasks that would mirror processes that were used in classrooms as well as tasks that resembled more traditional classroom tests with open-ended items. The model for the mathematics task used in this study included three stages:

1. Setting a context: Students look at stimulus materials that may include written text, graphs, tables, charts, or other graphics that present mathematical information.

2. Short Answer Items: Students respond to multiple short-answer items designed to have them analyze the information presented in the stimulus materials.

3. Integration Item: Students read a prompt and use the previous analyses to write an extended discussion (which could include predictions) of the information presented in the stimulus materials.

7. Self-reflection: For any graphic display, students use a checklist to evaluate the effectiveness of the display and to guide revisions, if needed.

The committee of teachers that developed the prototype tasks also insisted that introductory directions tell students what steps they would be completing during extended tasks, as well as the bases for evaluation before they began any performance task.

The two test forms began in exactly the same way. Students were given general directions to introduce the task and the bases of evaluation:

> The table below shows the number of high school students and the number of high school dropouts in Washington state for the years 1988 through 1991. The dropout rate for each of these years is also given. In addition, a graph illustrating the data is given for each data set. In the items that follow, you will look at the data in the table and the graphs and then describe the trends in enrollment, number of dropouts, and dropout rate from 1988 to 1991. You will also make predictions of what the numbers probably looked like in 1995 if the trends continued in the same fashion. You will be evaluated on your statistical understandings and the effectiveness of your mathematical communication.

These directions were provided in the student response books and were also read aloud to students so that reading skill would not detract from student performance. Students were then given three graphs and a table showing the school drop-out and enrollment data (See Figure 1).

insert Figure 1 about here

Next, students responded to six short-answer items designed to have them focus on the information presented in the table and graphs. The data showed a decrease in enrollments and drop-outs but an increase in drop-out rate. Students were asked to analyze each graph separately, discuss whether drop-out numbers could drop while drop-out rate rose, identify the graph that would support an argument that the school drop-out rate was improving, and identify the graph that would support an argument that the school drop problem was getting worse. Items 1 through 6 are given to students are given in Figure 2.

insert Figure 2 about here

After analyzing the school drop-out data via items 1 through 6, students worked were asked to write a letter to the governor discussing their positions on the trends in school drop-out. The directions for the letter differed in the two test forms. In the "connections" condition, students were given the following prompt:

> Decide if you think the dropout problem is improving, remaining constant, or getting worse. Draft a brief letter to the Governor. In your letter:
>
> - state your position about the dropout problem
> - tell what you think 1995 drop-out data looks like given the trends in the data.
> - **use data from the table and graphs to support your position and prediction**
> - add your own knowledge
>   - if you think the drop-out problem is getting worse, give one possible cause for the problem
>   - if you think the drop-out problem is getting better, give one possible reason for the improvement
>   - conclude your letter with a recommendation about how to eliminate school dropout

For the "no connections" condition, students were given the following prompt:

> Decide if you think the dropout problem is improving, remaining constant, or getting worse. Draft a brief letter to the Governor. In your letter:
>
> - state your position about the dropout problem
> - tell what you think 1995 drop-out data looks like given the trends in the data.
> - **use data from the table and graphs to support your position and prediction**

As can be seen, the prompts were identical except for the directions to add their own ideas regarding the school drop-out problem and a recommendation on how to eliminate it.

Scoring Methods. Two methods were used to score the mathematics tasks: "trait" scoring and item by item scoring. For the first scoring method, "trait" scoring, students received two holistic scores: one for understanding of statistics and probability concepts and one for mathematical communication. The scoring criteria were developed by a committee of mathematics teachers and refined using student work from the pilot assessment program. The scoring criteria were each on a four point scale (see Figure 3 for the scoring criteria for statistics and probability). Raters applied each of the trait scoring rubrics to students' responses to the task as a whole.

---

insert Figure 3 about here

The second scoring method involved item-by-item scoring. Each of the six statistical analysis items were scored using a 2 or 3 point scoring rule and the letter (item 7) was scored with a 4 point scoring rule. Scoring rules for individual items took into account both accuracy of conceptual understanding and completeness of a response.

1. Item 1 was scored using a 2 point scoring rule (1 or 0). To earn 1 point, students had to accurately describe the decrease in enrollment shown in the table of data and Graph 1.

2. Items 2 and 3 were scored using a 3 point scoring rule (2, 1, or 0). To earn 2 points, students had to accurately describe the trend evident in the table and relevant graph and use accurate data from the graph or table to support their response. To earn 1 point, students had to accurately describe the trend but provided no supporting data or inaccurate data.

3. Item 4 was also scored using a 3 point scoring rule (2. 1. or 0). To earn 2 points, students had to describe that the trends in dropout numbers and dropout ratios could be different due to a faster drop in enrollment numbers than in drop-out numbers. To earn 1 point, students could either make an incomplete description of that relationship or show some confusion about the relationship, while still acknowledging that the enrollment numbers were a factor.

4. Items 5 and 6 were also scored using a 3 point scoring rule (2, 1, or 0). To earn 2 points, students had to identify the correct graph and tell what the graph presented that supported the argument.

5. Item 7, the letter to the Governor, was scored using a 4 point rubric focused on both mathematical understandings shown in the task (e.g., a reasonable prediction based on evidence from the graphs and table) and clear communication of the mathematical ideas. The directions for the letter indicated that students were to use data from the graphs or table in their responses. In order to receive all 4 points, students had to communicate their positions clearly using data to support their positions and make a prediction based on their positions, including data to support the prediction. To receive 3 points, students had to communicate their positions and predictions and support at least one of these points with data. To receive 2 points, students had to communicate their positions and/or predictions but typically did not include data in support of the points. Finally, students who earned 1 point took a position or make a prediction but the majority of the response was based on ideas about the causes of school drop-out or potential solutions to the problem.

Procedure

Materials were sent to the teachers in the spring of 1995. Teachers were asked to administer the tasks and return materials in the postage-paid return envelopes. All materials were received between March 31 and May 31, 1995.

Four research assistants were hired to score the mathematics tasks. Pairs of raters were assigned to one of the two scoring conditions. One pair scored the tasks using item by item scoring procedures and one pair scored the tasks using the two trait rubrics (understanding of statistics and probability concepts and mathematics communication). Pairs of raters were trained separately so that they could focus only on the scoring rules they were to apply to the students' work. Raters were blind to the purpose of the study.

All raters participated in a 1 to 2 hour training session. Prior to beginning the scoring process, raters completed the "no connections" version of the task themselves. They discussed the task and what was demanded by it.

The pair that used item-by-item scoring rules reviewed and discussed the rules and then scored 8 anchor papers independently (anchor papers represented a range of student work). They discussed their scoring decisions with the researcher and each other, and worked toward a consensus agreement on scores. They scored another 8 anchor papers, reaching even closer agreement, and worked toward a consensus on scores. Then they scored the remaining 160 papers in the research set. Rater agreement for item-by-item scoring on the remaining 160 papers was 100 percent for adjacent scores on the six statistics items and the letter. Exact score agreement was 91 percent for the six statistics items and 78 percent for the letter.

The pair of raters who applied the trait scoring rules reviewed the scoring criteria for the statistics and probability trait and discussed its meaning, including how the trait was distinct from other relevant dimensions of mathematics and how the scoring rules could be applied to make a judgment across all responses in the task. Raters then scored the first set of 8 anchor papers. Once they had completed scoring the 8 papers, raters met with the researcher and discussed their ratings. For all papers, the scores for each rater were within one point of each other. Raters discussed all 8 papers, attending to points of agreement as well as disagreement, and then assigned a final score to each of the 8 papers.

Raters then scored the second set of anchor papers and repeated the discussion process. Raters scored the remaining 160 papers independently. Once raters completed scoring for understanding of statistics and probability concepts, they repeated the process for the mathematical communication trait. Rater agreement for adjacent score matches for the two trait scores ranged from 99 to 100 percent. Rater agreement for exact matches on the two traits was 66 percent for statistics and probability scores to 83 percent for mathematical communication scores.

For both scoring methods, papers for which there was not agreement were scored by a third reader. Consensus between the third reader and one of the original raters established the final scores for the item or task. Once item scores were established for the six statistics concepts items (using the item by item scoring rules), a total score was obtained for statistical concepts by summing scores across the six items. The resulting scores used in the analyses were:

1. Trait scores: mathematical communication and understanding of statistical and probability concepts and procedures.

2. Item-by-item scores: statistics total and letter score.

Results

Group differences were investigated in this study separately for the two scoring methods. It was anticipated that there would be no differences in scores between the students in the "connections" condition and students in the "no connections" condition for either scoring method. Two types of analyses were used to assess the data. First of all, homogeneity of variance tests were used to determine whether variability of scores differed for the two groups using either of the scoring methods. Secondly, two one-way ANOVAs were conducted for each scoring method to determine whether mean scores for the groups differed on the scores.

Trait Scoring. Table 2 presents the number of students in each condition earning each score level for each of the mathematical traits. As can be seen, the majority of students in both conditions earned a score of 2 on both traits (46 to 61 percent). About one-third of the students in both conditions earned scores of 3 for both traits. Table 3 presents the means, standard deviations, and variances for each of the mathematical traits for each group. Cochran's homogeneity test results are presented in Table 4 (df = 87,2). The variances for the students in the two conditions were not significantly different for either trait score. While not significant, the students in the "no connections" condition had greater score variability than did the students in the "connections" condition for mathematical communication.

Two one-way ANOVAs were conducted; one for each mathematics trait. Table 5 presents the results of the ANOVAs. For neither of the traits were the means significantly different. Given the score distributions shown in Table 2 it appears that neither group was very proficient in communicating mathematical ideas or in using statistical concepts and procedures effectively.

<u>Item by Item Scores</u>. Table 6 presents the number of students in each condition earning each score level for the letter to the Governor. As can been seen, there is a trend toward higher scores for the "no connections" group than for the "connections" group. Table 7 presents the means, standard deviations, and variances for the statistics total score and the letter score for each group. Cochran's homogeneity test results are presented in Table 8 (df = 87,2). The variances for the students in the two conditions were not significantly different for either score.

Two one-way ANOVAs were conducted; one for the statistics total score and one for the letter score. Table 9 presents the results of the ANOVAs. While the two groups were not different in terms of statistical understandings, the mean scores for the letter were significantly different ($F_{1,168} = 7.705$, $p < .01$). In fact, the mean scores for the two groups were different by .5 out of 4 points for the letter ("No connections" mean = 2.57 and "connections" mean = 2.04), while the mean scores for the two groups on the statistical total score were only different by .04 out of 11 possible points. Clearly the two groups were equivalent in terms of their understanding of statistics and probability concepts but were quite different in their communication of mathematical ideas through the letters.

Looking at the specific score points for the letter shown in Table 6, 17 percent of the students in the "no connections" condition earned a score of 4, indicating that, in responding to the prompt, they had included data from the table or graphs to support their positions and predictions. In contrast, only 7 percent of the students in the connections condition earned a score of 4. In addition, 22 percent of the students in the "connections" condition earned a score of 1 for the letter, indicating that they focused mostly on the reasons for school drop-out and ways to solve the problem rather than providing mathematical information to support their

positions and/or predictions. Another 7 percent of the students in this condition earned scores of 0, indicating that they either did not address the prompt or neither took a position nor made a prediction about school drop-out (missing data were omitted from the analyses). In contrast, only 11 percent of the students in the "no connections" condition earned a score of 1 and none of the students earned a score of zero. The majority of students in both conditions earned scores of either 2 or 3, indicating that they attempted to use mathematics to support their positions and either omitted data or only included data to support either the prediction or the position on trends. These data suggest that, for this context, one in which students are presented with data related to an issue that may be of importance to them and their own lives (i.e., school drop-out), students may have been distracted from the mathematical requirements of the task by requests for personal ideas and opinions.

Discussion and Conclusion

This study provides a view of the impact of prompt structure on students' scores when students are asked to support their views with both mathematical information and personal ideas and opinions. Specifically, students were asked to write a letter in which they were required to take a position on the trends in school drop-out, support that position with evidence from tables and graphs, predict what would happen in four years based on the trends, and support that prediction with data. For one condition, students were also asked to add their own ideas about the causes of trends in school drop-out and ways to solve the problem. The ANOVAs showed that, when item-by-item scoring methods were used, mean scores on the letters were lower for students who were asked to add their own ideas than for students who were not asked to add their own ideas.

This study was limited, however, in that it investigated the impact of prompt structure on a *single* topic. The topic may have been one that triggered strong emotional responses among the students. In order to further explore the question of impact of prompt structure on examinee performance, additional studies are needed. A variety of prompts that ask students to write about many, potentially controversial topics as well as topics of a less controversial

nature should be tested. Meanwhile, test developers should use caution when developing tasks that ask students to make connections between information within the task and their own knowledge and ideas.

In addition, scoring procedures used can have an impact on students' scores. These results indicate that holistic trait scores for mathematical communication were not different when scores were based on the students' responses to the task as a whole. It is possible that students did a fairly decent job of communicating mathematical ideas in the remainder of the task so the letter carried less weight when an overall trait score was used instead of focusing the communication score on a single item.

From the point of view of instructional planning, more attention needs to be paid to teaching students how to make arguments and conjectures using mathematical information to support their positions. If students are to be held accountable for their mathematical communication, they will need to learn how to make cases using evidence rather than emotion. Figures 4a and 4b show examples of two letters: one from a student who earned a score of 4 for the letter and one from a student who earned a score of 1. Both letters are from students in the "connections" condition. Clearly it is possible to make a strong case for a position and/or prediction, support it with evidence, and still make strong points about the causes for trends and possible outcomes. If we hope to engage students in the assessment process, we need to continue to explore topics of interest to students. On the other hand, if students are to be deemed proficient in mathematical communication, we also need to ensure that they are taught how to communicate their ideas and opinions effectively.

insert Figures 4a and 4b about here

The results found here may be due to the age of the students, the nature of the prompt, or the topic of the task. As states and districts attempt to develop mathematics assessments that allow more time for thoughtful and reflective analysis of mathematical information, research is needed about the most effective ways to frame assessment tasks. As long as large scale testing

programs are used to make decisions about students, schools, and districts, continued research on how to frame the assessments given to examinees merits attention. In addition, with the increase in the use of performance-based assessments, studies about wording of prompts, task structure, and scoring methods are clearly needed. Finally, as teachers we must focus our teaching on helping our students to be effective mathematical communicators, giving them examples of effective communication - even about controversial issues - and giving them practice with the skills inherent in the effective communication of mathematical ideas.

References

Baxter, G. P., Shavelson, R. J., Herman, S. J., Brown, K. A., & Valdez J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact. *Journal for Research in Mathematics Education, 24* , 190-216.

Brossell, G. & Ash, B. H. (1984). An experiment with the wording of essay topics. *College Composition and Communication, 35*, 423-425.

Brossell, G. (1983). Rhetorical specification in essay examination topics. *College English, 45*, 165-173.

Cherry, R. D. (1989) . Fictional scenarios and rhetorical specification in writing tasks: A cautionary note. *Journal of Advanced Composition, 9* (1-2), 151-161.

Goldberg, G. L., & Kapinus, B. Problematic responses to reading performance assessment tasks: Sources and implications. *Applied Measurement in Education, 6*, 281-305.

Hoetker, J. (1982). Essay examination topics and students' writing. *College Composition and Communication, 33*, 377-392.

Hoetker, J., & Brossell, G. (1986). A procedure for writing content-fair essay examination topics for large-scale writing assessments. *College Composition and Communication, 37*, 328-335.

Hoetker, J., & Brossell, G. (1989). The effects of systematic variations in essay topics on the writing performance of college freshmen. *College Composition and Communication, 40*, 414-421.

National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics* . Reston, VA: The Council, 1989.

Ruth, L., & Murphy, S. (1984). Designing topics for writing assessment: Problems of meaning. *College Composition and Communication, 35*, 410-422.

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49* (8), 20-25.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: political rhetoric and measurement reality. *Educational Researcher, 21* (4), 22-27.

Table 1

Composition of the Sample for the Mathematics Task

Condition	Number of Teachers	Number of Students	Grade Level	Percent of Students at Grade Level
No Connections Requested	3	36	10	100%
Connections Requested	5	140	10	11%
			11	89%

Table 2

Percent of Students at Each Score Level for Each Condition and for Each Mathematical Trait

		Percent at Each Score Level			
Mathematical Trait	Conditions	1	2	3	4
Statistics and Probability	No Connections Requested	8.3	52.8	33.3	5.6
	Connections Requested	12.1	45.7	33.6	8.6
Math Communication	No Connections Requested	11.1	47.2	38.9	2.8
	Connections Requested	6.4	60.7	32.1	0.7

Table 3

Means, Standard Deviations, and Variances for Each Group on Each Mathematical Trait

Trait	Condition	N	Mean	SD	σ^2
Statistics and Probability	No Connections Requested	36	2.36	.72	.52
	Connections Requested	140	2.38	.81	.66
Math Communication	No Connections Requested	36	2.33	.72	.51
	Connections Requested	140	2.27	.59	.34

Table 4

Results of Homogeneity of Variance Analysis for Mathematical Traits

Trait	Cochran's C	Approximate P
Statistics and Probability	.56	.29
Math Communication	.60	.06

Table 5

Univariate Analysis of Variance for Each Mathematical Trait

Trait	Source	SS	DF	MS	F	Sig.
Statistics and Probability	Main Effect	.017	1	.017	.028	.868
	Residual	109.48	174	.629		
Math Communication	Main Effect	.110	1	.110	.291	.590
	Residual	65.69	174	.378		

Table 6

Percent of Students at Each Score Level for the Letter for Each Condition

		Percent in Each Score Level				
Score	Condition	0	1	2	3	4
Letter	No Connections Requested		11.4	37.1	34.3	17.1
	Connections Requested	6.7	20.7	40.0	25.2	7.4

Table 7

Means, Standard Deviations, and Variances for Each Group on Statistics Total Score and the Letter Score

Score	Condition	N	Mean	SD	σ^2
Statistics Total	No Connections Requested	36	9.36	1.48	2.18
	Connections Requested	140	9.32	1.35	1.83
Letter	No Connections Requested	35	2.57	.92	.84
	Connections Requested	135	2.06	1.01	1.03

Table 8

Results of Homogeneity of Variance Analysis for Statistics Total Score and the Letter Score

Score	Cochran's C	Approximate P
Statistics Total	.55	.324
Letter	.55	.361

Table 9

Univariate Analysis of Variance for Statistics Total Score and the Letter Score

Trait	Source	SS	DF	MS	F	Sig.
Statistics Total	Main Effect	.03	1	.03	.016	.900
	Residual	331.191	174	1.903		
Letter	Main Effect	7.291	1	7.291	7.374	.007*
	Residual	166.097	168	1.026		

* significant at < .01 level

Figure 1

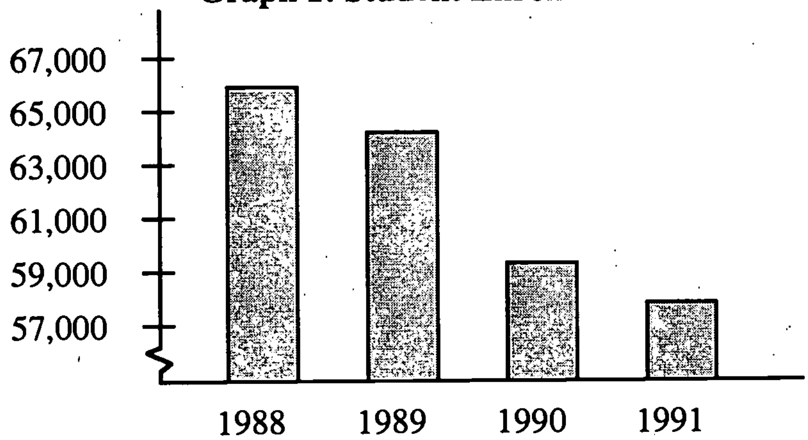
Data and graphs given to students for the mathematics task.

Use the table and graphs to do Numbers 1 through 7. You may refer back to the table and graphs as often as you need to.

Washington State High School Enrollment Data

	1988	1989	1990	1991
Student Enrollment	65,920	64,260	59,280	57,760
Dropouts	16,060	15,850	15,580	15,390
Dropout Rate	0.24	0.25	0.26	0.27

Graph 1: Student Enrollment



Graph 2: Number of Dropouts

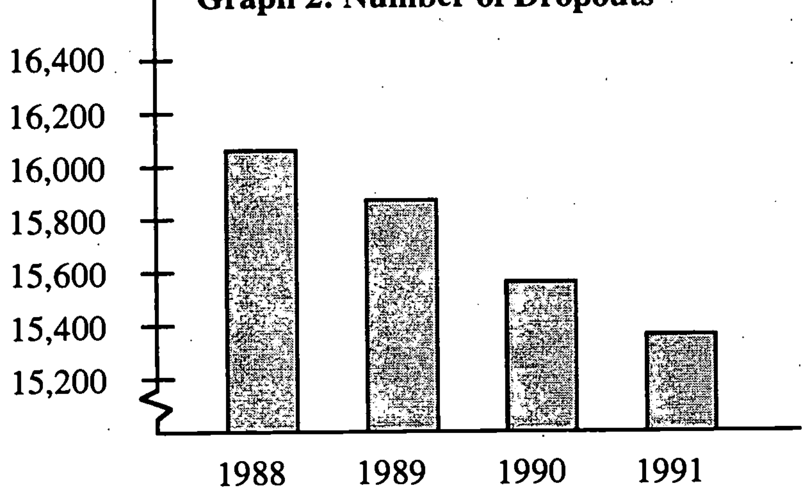


Figure 1 (Continued)

Data and graphs given to students for the mathematics task.

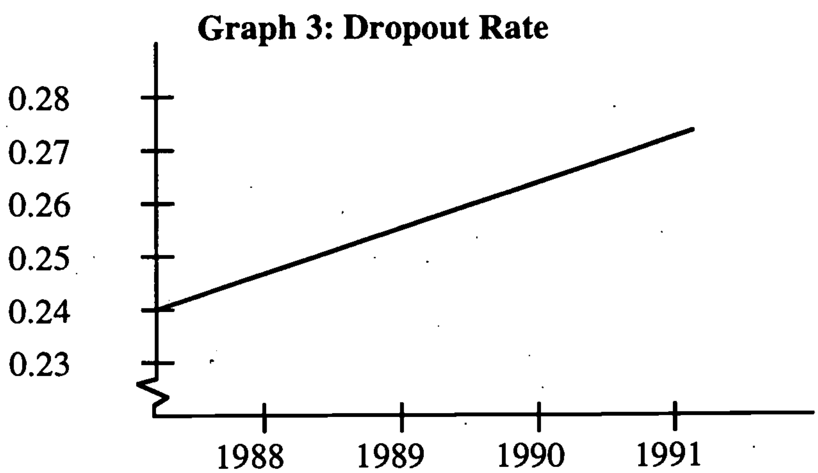


Figure 2

Statistical analysis items for the school drop-out task.

1. Use the table and Graph 1: Student Enrollment to describe the trend in high school enrollment from 1988 through 1991.

2. Use the table and Graph 2: Number of Dropouts to describe the trend in high school dropouts from 1988 through 1991. Use specific data from the graph or table in your answer.

3. Use the table and Graph 3: Dropout Rate to describe the trend in the Washington high school dropout rate from 1988 through 1991. Use specific data from the graph or table in your answer.

4. Compare the trends in number of dropouts with the dropout rate. Is this situation possible? If it is, explain how it could occur. If it is not, explain why. Refer to data from the table or graphs in your response.

5. Chris argues that the dropout problem is improving. Which graph supports Chris's argument? Tell why.

6. Terry argues that the dropout problem is getting worse. Which graph supports Terry's argument? Tell why.

Figure 3

Trait Scoring Rubric for Statistics and Probability Concepts and Procedures

Mathematics Scoring Criteria

Probability and Statistics Performance Criteria

Chance:	understands concepts of chance (certainty and uncertainty, experimentation and theory, probability, dependence and independence
Data Analysis:	understands concepts of data collection and analysis (population and sampling, central tendency and distribution)
	conducts data analyses (collects data, analyzes central tendency and distribution, displays results in tables, graphs, and charts)
	understands how to interpret data (inference, point of view, uses and misuses)

SCORING

4 points Meets or exceeds all relevant criteria
- shows **extensive** understanding of concepts and procedures both within and beyond the task
- **consistently and purposefully** applies appropriate concepts and procedures

3 points Meets all relevant criteria
- shows **thorough** understanding of concepts and procedures required by the task
- **consistently** applies appropriate concepts and procedures

2 points Meets some relevant criteria
- shows **general** understanding of concepts and procedures required by the task
- **generally** applies appropriate concepts and procedures

1 point Meets few relevant criteria
- shows **rote or partial** understanding of concepts and procedures required by the task
- **occasionally** applies appropriate concepts and procedures

Figure 4a

Example of level 4 response for letter to the Governor

- Given the data from 1988 - 1991, I think the dropout problem is staying fairly constant. The actual number of dropouts only changed by 670 students. This is no big deal when you see that 8,160 less students were enrolled.
- Given the trends from 1988 - 1991, it would appear that there would be about 47013 students enrolled in 1995, and the dropout rate would be about .31, about 14,574 dropouts. However, there may be a change in the trend that I'm not aware of, since the data did stop in 1991. It could have turned around.
- Since I think the problem is staying constant, I have no answer for this part.

- You can't eliminate school dropout. There will always be someone who is convinced they don't need an education, and they arent going to listen to anyone else. You could decrease dropout by stressing the importance of an education. Go around to elementary schools, show them statistics. Things like that.

STOP

BEST COPY AVAILABLE

Figure 4b

Example of level 1 response for letter to the Governor

Governor ~
There is an increasing number of dropouts
and it needs to Stop because the
dropouts go on the streets & get hurt.
Our generation is going to Kill itself,
and droping out is a part of it.
By 1995 I think there will be so
many dropouts there won't be very
many left in the schools. Like in
Graph 3 it shows this. One possible
cause for more dropouts is because
you can go out & get a job just
as easy as people who go
to school. Actually it's easier
because they can work more
because they don't go to school.
It is necessary to have it
where School is more interesting
STOP
So people don't want to dropout.

TM026431
NCME Annual Meeting, April 9-11, 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Influence of Task Directions on Student Performance for Open-Ended Mathematics Assessments.	
Author(s): Catherine S. Taylor	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] **Sample sticker to be affixed to document**

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here
Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Catherine S. Taylor	Position: Assistant Professor
Printed Name: Catherine S. Taylor	Organization: University of Washington
Address: 312 Miller Hall, Box 353600 College of Education Seattle, WA 98195-3600	Telephone Number: (206) 543-1139
	Date: 2/4/97

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

March 12, 1996

Dear NCME Presenter,

Congratulations on being a presenter at NCME[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the NCME Conference. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (23)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: NCME 1996/ERIC Acquisitions
O'Boyle Hall, Room 210
The Catholic University of America
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the NCME web page (http://www.assessment.iupui.edu/ncme/ncme.html). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an NCME chair or discussant, please save this form for future use.